NATURAL ASSISTANT INTERACTION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Application Ser. No. 62/648,084, entitled "NATURAL ASSISTANT INTERACTION," filed on Mar. 26, 2018, the content of which is incorporated by reference in its entirety for all purposes.

FIELD

[0002] This relates generally to virtual assistants and, more specifically, to providing natural language interaction by virtual assistants.

BACKGROUND

[0003] Virtual assistants (or digital assistants or intelligent automated assistants) can provide a beneficial human-machine interface. Such assistants can allow users to interact with devices or systems using natural language in spoken and/or text forms. For example, a user can provide a speech input containing a user request to a digital assistant operating on an electronic device. The virtual assistant can interpret the user's intent from the speech input and operationalize the user's intent into tasks. The tasks can then be performed by executing one or more services of the electronic device, and a relevant output responsive to the user request can be returned to the user.

[0004] Virtual assistants can be activated upon receiving a trigger phrase such as "Hey Siri." Upon activation, virtual assistants can receive and process user's speech input. For example, a user's speech input may include a leading trigger phrase to activate the virtual assistant followed by a request for information (e.g., "Hey Siri, how is the weather today?"). Leading every speech input with a trigger phrase (e.g., "Hey Siri"), however, can be inconvenient and quickly become cumbersome. It also does not represent a natural way of communication. For example, when a first user talks to a second user, the first user typically would not lead every sentence with the name of the second user. Thus, requiring the user to lead each speech input with a trigger phrase does not represent a natural way of communication and is less efficient.

SUMMARY

[0005] Systems and processes for providing natural language interaction by a virtual assistant are provided.

[0006] In accordance with one or more examples, a method includes, at an electronic device with one or more processors, memory, and a microphone: receiving, via the microphone, a first audio stream including one or more utterances and determining whether the first audio stream includes a lexical trigger. In accordance with a determination that the first audio stream includes the lexical trigger, the method further includes generating one or more candidate text representations of the one or more utterances and determining whether at least one candidate text representation of the one or more candidate text representations is to be disregarded by the virtual assistant. In accordance with a determination that at least one candidate text representation is to be disregarded by the virtual assistant, the method further includes generating one or more candidate intents based on candidate text representations of the one or more candidate text representations other than the to be disregarded at least one candidate text representation. The method further includes determining whether the one or more candidate intents include at least one actionable intent. In accordance with a determination that the one or more candidate intents include at least one actionable intent, the method further includes executing the at least one actionable intent and outputting a result of the execution of the at least one actionable intent.

[0007] Example non-transitory computer-readable media are disclosed herein. An example non-transitory computer-readable storage medium stores one or more programs. The one or more programs comprise instructions, which when executed by one or more processors of an electronic device, cause the electronic device to receive, via a microphone, a first audio stream including one or more utterances; determine whether the first audio stream includes a lexical trigger; in accordance with a determination that the first audio stream includes the lexical trigger, generate one or more candidate text representations of the one or more utterances; determine whether at least one candidate text representation of the one or more candidate text representations is to be disregarded by the virtual assistant; in accordance with a determination that at least one candidate text representation is to be disregarded by the virtual assistant, generate one or more candidate intents based on candidate text representations of the one or more candidate text representations other than the to be disregarded at least one candidate text representation; determine whether the one or more candidate intents include at least one actionable intent; in accordance with a determination that the one or more candidate intents include at least one actionable intent, execute the at least one actionable intent; and output a result of the execution of the at least one actionable intent.

[0008] Example electronic devices are disclosed herein. An example electronic device comprises one or more processors; a memory; and one or more programs, where the one or more programs are stored in the memory and configured to be executed by the one or more processors, the one or more programs including instructions for, receiving, via the microphone, a first audio stream including one or more utterances; determining whether the first audio stream includes a lexical trigger; in accordance with a determination that the first audio stream includes the lexical trigger, generating one or more candidate text representations of the one or more utterances; determining whether at least one candidate text representation of the one or more candidate text representations is to be disregarded by the virtual assistant; in accordance with a determination that at least one candidate text representation is to be disregarded by the virtual assistant, generating one or more candidate intents based on candidate text representations of the one or more candidate text representations other than the to be disregarded at least one candidate text representation; determining whether the one or more candidate intents include at least one actionable intent; in accordance with a determination that the one or more candidate intents include at least one actionable intent, executing the at least one actionable intent; outputting a result of the execution of the at least one actionable intent.

[0009] An example electronic device comprises means for receiving, via the microphone, a first audio stream including one or more utterances; means for determining whether the first audio stream includes a lexical trigger; in accordance